DOCUMENT RESUME

ED 172 807 IR 007 442

AUTHOR Lewis, Richard F.
TITLE The Formative Evaluation of Bilingual Television Programs: Some Results and Suggestions for Improvement.
PUB DATE Jan 79
NOTE 36p.; Paper presented at the Annual Convention of the Association for Educational Communications and Technology (New Orleans, Louisiana, March 1979); For related document, see IR 007 416

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Attitudes; *Educational Television; *Evaluation Criteria; Evaluation Methods; *Formative Evaluation; French; Grade 8; *Instructional Aids; Interaction; *Language Instruction; *Second Languages

ABSTRACT

This paper concerned with the improvement of formative evaluation procedures for second language television materials includes a literature review and descriptions of two studies of French second language teaching materials produced by Nova Scotia School Television. Evaluation procedures used in determining vocabulary acquisition, program enjoyment, and plot comprehension for the first series, Vive La Compagnie, are explained; the second series, Allons-Y, was evaluated in four studies to determine the effects of program viewing without teacher input, with teacher input, between group interaction, and attitudes toward the program. Evaluation criteria relating to language materials are defined and listed in three basic categories: those factors which can be examined independently, those requiring viewing the program, and those which need viewer interaction with the program. A list of references is attached. (RAO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED172807

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Formative Evaluation of Bilingual Television Programs: Some Results and Suggestions for Improvement

Richard F. Lewis, Ph.D.
Research Associate
Atlantic Institute of Education

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Michael Simonson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

January, 1979

Paper presented at the Association for Educational Communication and Technology meeting New Orleans, La. March, 1979

I 800 7442

650

ABSTRACT

The paper presents the results of two formative evaluation studies of second language programs in Nova Scotia. Both studies included procedures to determine the effectiveness of the series in achieving their objectives as well as procedures which could yield suggestions for improvement in the programs.

A second aspect of the paper concerns the improvement of formative evaluation procedures for television and other mediated programs. The suggested changes are divided into factors which can be examined independently, factors which must be examined in context and factors which must be examined in a field tryout.

The educational media have a definite role to play in helping children and adults to learn other languages. Radio, audio-tapes, films and slide sound packages and television have all been used to achieve various learning objectives. Many of the early language media presentations consisted of stand up teaching using blackboards and perhaps a few visuals but little else. However, with the maturation of educational television, more sophisticated presentations are being designed to teach languages.

Many agencies have produced series for use in second language teaching. The Ontario Educational Communications Authority in Toronto, has produced several series for use in teaching French while the Carrascolendas series has been used to teach Spanish in the United States. Nova Scotia School Television also produced several series to supplement teaching of French as a second language.

Most of the recent television series intended for second language learners have been subjected to formative evaluation or at least summative evaluation procedures. These studies have dealt with various aspects of the production and utilization of second language television materials. (See Laosa, 1975; Williams, Van Wart and Stanford, 1973; Fleming, 1972)

This paper deals with the improvement of formative evaluation procedures for second language television materials. The recommended procedures are based on a review of the literature and on two studies of second language teaching materials produced by Nova Scotia School Television (Lewis and Fisher, 1977, and Lewis, 1979). Although the paper deals with procedures for television programs, the comments could be applied to other media and other programs.

The Nova Scotia Studies

The recommendations on improving procedures are based on two studies of second language television programs. To provide the reader with this background, these studies are briefly summarized below.

Vive La Compagnie

Vive La Compagnie was a second language French series designed for grade eight students. In each program, three actors (and other supporting actors) participated in real-life experiences which might be faced by students. The programs had a controlled vocabulary and relied on humor and visuals to convey meaning.

The Evaluation

The evaluation was commissioned by the Nova Scotia Department of Education to determine the effectiveness of the program series achieving its objectives, which dealt with vocabulary acquisition, enjoyment and understanding of the plots. In addition to determining effectiveness, the study had to provide specific information on recommended production changes, distribution patterns, usage guidelines and supplementary materials.

Specific Questions

1. Would the students watch the programs; i.e. would their eyes be on the screen?
2. Would the students understand the words to be used in each program?
3. After watching the programs, would students understand words which were not in their repertoire before the viewing?
4. Would students understand the plots?
5. Would students enjoy the programs?

6. Would the number of programs watched result in differences in vocabulary acquisition, understanding and enjoyment?

7. How would students react to selected variables such as the plot, the characters and production techniques?

8. Would the students' attitudes to learning French and towards French people be altered by exposure to the program?

9. How would the teachers feel about the program with regard to its level, its achievement of objectives, etc.

Data Sources

The evaluation study used three basic sources of data: expert opinion, student opinion and test results, and teacher opinion. The expert opinion consisted of a contextual evaluation in which the programs were scrutinized to determine the production variables which could be improved. The actions of the characters, the pronunciation and the special effects all came under close scrutiny for suitability. In addition, the relationship of the production format, etc., to existing research recommendations for children's programs was investigated. The researcher determined whether the recommendations of experts in children's programming were being considered in the Series.

Teacher attitude was obtained using a Likert type scale similar to the students' scale. In addition, each participating teacher was interviewed to determine his/her reaction to every aspect of the program which could, if necessary, be changed or altered.

Design, Subjects, Procedure

Two hundred and eighty-eight grade eight students from 12 classes participated in the study. These students came from four different schools.

Classes were chosen at random from all the grade eight French classes in one Nova Scotia county. To determine the effects of watching more programs, some classes watched eight programs, some watched four and some watched one program; other classes did not watch any programs but participated in the study so that valid comparisons could be made. All classes completed a baseline test battery before the study began while one-half of the sample completed a pre-test. Classes viewing the program watched it on videotape in their classrooms during the regular French class. Discussion of a program followed its showing. After watching the appropriate number of programs, all the classes were tested on the same program using the techniques mentioned above. To obtain the most representative sample of teacher opinion regarding the program, any teacher using the program outside the study as well as several other groups of teachers completed the teacher questionnaire on the series.

Pre-Treatment Tests for Differences

In this study, classes had to be randomized, not students. As a result, differences between classes had to be measured before the treatment was presented. Two tests were administered to the sample prior to the study.

The Ici on Parle Test, an audio-pictorial vocabulary test, discovered significant differences between groups. In every case, means for the TV group appeared to be higher than means for the non-TV group.

Significant differences between groups were also found in responses to the questionnaire on attitude to learning French. This test comprised two sub-tests, Desire to Learn French and Motivational Intensity.

Results*

Attention

The programs maintained the attention of students watching about 80% of the time. However, certain sections of programs did seem to encourage students to stop watching programs. In addition, it appeared that students who did not understand the action (determined by self-report) stopped watching to catch up on their sleep, homework, etc. Sections of programs which were accompanied by lapses in attention were noted for further analysis.

Vocabulary Level Before Viewing

The 144 students who completed the pre-test knew 14 of the 20 words on the test. This data indicated that students could understand the major words and sentence expression used in the program. This finding showed that the assessments by the teachers and of the producers regarding vocabulary level were reasonably accurate.

Vocabulary Level After the Program

The difference between pre-test and post-test vocabulary scores indicated that students did not learn new words or phrases as a result of the program. Neither the pre-test group, nor the non pre-test group who watched the television showed a significant difference on the post-test from the non-television groups. However, it should be pointed out that the program was not designed to teach vocabulary.

* This section contains only summary information since the paper is concerned mainly with the improvement of procedures. Complete results are presented in Lewis and Fisher (1977).

The data did indicate that a student who had been exposed to a pre-test would score significantly higher than a student who was not exposed to the pre-test. As a result, one of the recommendations suggested that as an instructional device, students be exposed to the vocabulary of the program before it was viewed.

Comprehension of Plots

Students who watched more programs understood more of the action than students who watched fewer programs. It is obvious that comprehension improved as the students became accustomed to the accent and pronunciation of the characters in the Series. These findings clearly demonstrated that viewing the whole series would be more beneficial than the viewing of just one program, or even just a part of the series.

Opinion of the Program

No significant differences were recorded between students who watched one, four, or eight programs. In ranking the programs, students chose programs which they said were easy to understand, funny, presented pictures to explain the words and had a gool plot. More information on these factors was sought and obtained in the group and personal interviews after each viewing and at the end of the study. The findings of the interviews and the summary of the questionnaires provided the basis for many of the production changes which were suggested.

Opinions of Teachers

Most teachers indicated that students would like the program but that a considerable amount of preparatory and follow-up work would be needed. Most teachers felt that the students would not understand the

vocabulary presented--an opinion which was completely contradicted by the actual performance of the students. A strong feeling of the teachers was that the pronunciation of one of the characters would inhibit learning the correct pronunciation. As a result, a change in characters or an improvement in pronunciation were suggested recommendations.

Attitude

The viewing of the program made no difference in the attitude scores of students. Those who watched did not attain higher attitude scores.

Problems with the Study

Experimenter Effect

In this study, the experimenter supervised the viewing of all programs and administered all tests. This procedure could hardly be called a natural situation for the classes involved. As a result, the generalizability of the data to a teacher administered situation could be called into question.

Vocabulary Testing

The same vocabulary test was used as the pre- and the post-test. The data indicated that the administration of the pre-test affected the post-test scores, i.e., students who took the pre-test socred significantly higher than students who did not complete the pre-test.

Attitude

Attitude is difficulty to measure. However, with the efforts of Gardner and Smythe (1975) attitude towards learning French can be measured. However, Gardner, Ginsberg and Smythe (1976) noted that attitudes of

post-tests seem to be lower than attitudes of pre-tests where there is no intervention (such as a television program). They attribute this finding to the fact that subjects tested are alerted to attitude which they did not have to express before.

Therefore, the measurement procedures used in this study may not be as valid as others which could be designed. The deletion of a pre-test of attitude with only post-test being administered may be a more desirable procedure. Alternatively, two equivalent forms of one instrument or two highly correlated instruments could be used.

Measurement of Attention

The attention measures were subject to a great deal of error since it was impossible to be certain whether a student was attending the program or just had his head turned in the direction of the television receiver.

ALLONS-Y

The Series consisted of ten fifteen minute programs. Each program contained a number of short segments which offered a range of topics. The range of segments included appearances by a teenaged host and hostess, singers, a band, game shows, activity programs and on-site visits to locations like Chéticamp, a fishing village in Cape Breton.

Supplementary Materials

Supplementary materials included a teacher's guide and a set of flashcards depicting key vocabulary and concepts used in the programs.

The teacher's guide included a list of the program's objectives, information on the segments presented in each program, a description of the

skit in each program and selected vocabulary and idiomatic expression. The guide, however, did not include any suggestions or directions for using the program.

Sixty-seven flashcards depicting vocabulary and concepts were available free of charge to teachers requesting them. These flashcards were printed in black and white on heavy cardboard.

Field Testing by Department Personnel

Three of the programs were viewed by students in 20 schools during visits by a staff member from the Education Resource Services. After the students viewed the program, they were asked to comment on the programs on an open ended questionnaire. Some of these comments are presented in a paper by Johnston-Doyle and Lewis (1978).

Purpose of the Evaluation

The purpose of the evaluation was to determine the effectiveness of the series in achieving its objectives and to make concrete recommendations for changes in the production and utilizations of the programs.

Specific Questions

1. Would viewing the program have an effect on students' ability to ___ comprehend words and sentences?
2. Would the program help students recognize correct constructions?
3. Would the students understand the content presented in the segments?
4. Would the students identify with the teen-aged host and hostess in the program?
5. Would the students like the musical segments presented in the program?

6. Would attitude to learning French be affected by viewing the program?
7. Would attitude to the French course be affected by viewing the program?
8. Would the addition of a teacher's preparatory and follow-up work affect scores on comprehension, ability to recognize constructions, attitude to learning French and attitude to the French class?
9. What changes in the program could be recommended as a result of the testing and the opinion of teacher and students.

The study was divided into four sub-studies in order to answer the variety of questions posed by the clients. The first study concerned itself with the effect of viewing the programs without any teacher input. The second study monitored the effects of the television program and teacher input. The third study dealt with the difference between the two groups in the first two studies. The final study surveyed attitudes towards the programs.

Sample

The studies used populations of grade eight students in Nova Scotia. In studies one and two, an experimental group/control group (Viewing/non-viewing) design was used with the same teachers teaching both a viewing group and a non-viewing group. In study one, the experimental group (78 students) watched the program without teacher input while the control group (92 students) continued with the specified curriculum. In study two, the 52 viewing students watched the program and participated in teacher preparatory and follow-up activity, while the 20 control group students continued with the curriculum. In study three, the 52 students who viewed

with the teacher were compared to the 78 students who did not have teacher input. Study four, which sought to measure attitudes and opinions, included 177 students who had viewed the programs.

Procedure

All students completed a baseline battery which included a test of French comprehension and vocabulary, an attitude test and a basic information questionnaire. Then, viewing groups watched eight programs while non-viewing groups proceeded with their regular work. After viewing the programs, all students completed the post-test battery. Viewing students completed a vocabulary test, a constructions test, a sentence comprehension test, a recall of skit test , three attitude tests and the tests on the opinion of each segment of the program. Non-viewing students completed only the vocabulary, sentence comprehension, constructions and three attitude tests.

Analysis

The analysis of covariance procedure developed by the Division of Educational Research Services at the University of Alberta was used to process the data. The analysis is computed on the basis of a pooled regression equation, homogeneity of regression is assumed. This procedure was used for studies 1, 2, and 3.

Results

Study 1

The purpose of study 1 was to determine whether exposure to the Series would result in an increase in post-test scores on ability and attitude.

Baseline Measures

The two groups (non-viewing and viewing) were not significantly different on the baseline measures. (See table 1). On the Ici on Parle Test, a test of French ability, the viewing group scored 25.31 while the control group scored 23.65. Both groups scored between 60 and 70 percent on the test. On the attitude test the viewing group scored 60.24 while the non-viewing group scored 56.53.

Post-test Measures

The scores of the two groups were significantly different only on the vocabulary test, after adjustment for the effect of the covariates measured in the baseline test. (See table 1). The groups were not significantly different on the sentence comprehension test, the construction test and the three attitude measures.

Table 1

Unadjusted, adjusted means and F. Ratios
for the viewing and non-viewing groups on all tests

	Unadjusted Means				Adjusted Means			
Test	Viewing	Non-Viewing	F. Ratio	df	Viewing	Non-Viewing	F. Ratio	df
Ici on Parle Test (Max=40)	25.31	23.65	2.62	1/168				
Attitude Test 1 (Max=90)	60.24	56.53	3.07	1/168				
Sentence Comprehension Test (Max=12)	4.74	4.18	3.84*	1/168	4.57	4.33	.99	1/166
Vocabulary Test (Max=16)	5.97	4.43	8.26*	1/168	5.59	4.76	5.20*	1/166
Constructions Test (Max=14)	6.06	6.20	.15	1/168	5.86	6.38	2.80	1/166
Attitude Test 2 (Max=61)	42.49	38.75	6.87*	1/168	41.66	39.44	3.21	1/166
Attitude Test 3 (Max=155)	90.17	87.18	1.37	1/168	90.46	86.94	1.85	1/166
Attitude Test 4 (Max=180)	104.18	97.46	4.05*	1/168	104.09	97.53	3.74	1/166

* p>.05

Study 2

The purpose of study 2 was to determine whether exposure to the Series would result in an increase in post-test scores on ability and attitude. However, in this study, the teacher actively worked with the class before and after the viewing, reinforcing the content.

Baseline Batteries

The viewing group scored significantly higher than the non-viewing group on the Ici on Parle test. (See table 2). The mean of the viewing group (31.85) was four points higher than that of the non-viewing group (27.85).

Post-tests

The scores of the two groups were significantly different on two tests, the vocabulary test and the sentence comprehension test (See table 2). The scores on the other tests were not significantly different, after adjusting for the covariates.

Table 2

Unadjusted, adjusted means and F. Ratios
for the Viewing and non-viewing groups on all tests

	Unadjusted Means				Adjusted Means			
Test	Viewing	Non-Viewing	F. Ratio	df	Viewing	Non-Viewing	F. Ratio	df
Ici on Parle Test (Max=40)	31.85	27.85	14.53*	1/70				
Attitude Test 1 (Max=90)	63.17	63.00	.28	1/70				
Sentence Comprehension Test (Max=12)	6.00	3.55	37.71*	1/70	5.87	3.88	24.85*	1/68
Vocabulary Test (Max=16)	11.77	4.10	81.79*	1/70	11.49	4.83	61.62*	1/68
Constructions Test (Max=14)	8.13	6.90	4.53*	1/70	7.93	7.44	.76	1/68
Attitude Test 2 (Max=61)	45.29	45.10	.00	1/70	45.69	[illegible]	.72	1/68
Attitude Test 3 (Max=155)	89.10	89.60	.06	1/70	88.96	89.93	.20	1/68
Attitude Test 4 (Max=180)	104.45	103.50	.14	1/70	104.63	103.26	.19	1/68

* p>.05

Study 3

The purpose of study 3 was to determine the effect of a teacher on the post-test scores. In the study, one group (TV) watched the program without any follow-up or preparatory activity while the other group (TVT) participated in activities in addition to viewing the program.

Baseline Test

The TV and Teacher (TVT) group scored significantly higher on the Ici Vocabulary Test than the TV only (TV) group. (See table 3). However, the two groups were not significantly different on the attitude battery. At 31.85, the mean for the TVT group was higher than the mean for the TV group.

Vocabulary Test

The TVT group was significantly different from the TV group. (See table 3). The adjusted scores were 6.51 for the TV group and 7.46 for the TVT group.

The efforts of the teacher obviously resulted in the increased vocabulary level of the group who had the teacher adding her input to the experience.

Sentence Comprehension Test

The adjusted scores for the two groups were not significantly different. (See table 3). The TVT group scored at 50% (6.0) before adjustment while the TV group scored at 30.5% (4.74).

The efforts of the teacher did not appear to significantly affect the results of the sentence comprehension test.

Constructions Test

The two groups were significantly different on the constructions test. (See table 3). When the scores were adjusted, the TVT group scored 9.99 while the TV group scores 7.16. Once again the teacher preparatory and following activities appeared to have a positive effect on the ability of the group to recognize correct constructions.

Attitude Tests

On all three attitude tests, the groups did not differ significantly, (See table 3), indicating that the intervention of a teacher along with the television series did not significantly affect students' opinions.

Table 3

Adjusted and unadjusted means for
TV only and TV and Teacher Groups on all variables

Test	Unadjusted Means				Adjusted Means			
	TV Only	TV and Teacher	F. Ratio	df	TV Only	TV and Teacher	F. Ratio	df
Ici on Parle Test (Max=40)	25.31	31.85	37.43*	1/128				
Attitude Test 1 (Max=90)	60.24	63.17	1.52	1/128				
Sentence Comprehension Test (Max=12)	4.74	6.00	14.98*	1/128	5.15	5.38	.57	1/126
Vocabulary Test (Max=16)	6.06	8.13	29.79*	1/128	6.51	7.46	6.86*	1/126
Constructions Test (Max=14)	5.97	11.77	70.39*	1/128	7.16	9.99	27.26*	1/126
Attitude Test 2 (Max=61)	42.49	45.29	3.28	1/128	43.19	44.23	.62	1/126
Attitude Test 3 (Max=155)	90.17	89.10	.85	1/128	89.79	89.67	.00	1/126
Attitude Test 4 (Max=180)	104.18	104.54	.04	1/128	103.44	105.65	1.13	1/126

* p>.05

Study 4

The purpose of study four was to obtain student opinions of the various production and content aspects of each of the segments. The overall findings seem to indicate that for the most part, with the exception of the Chéticamp segments, the students were able to follow the action and understand what was going on. Most of the segments were considered quite enjoyable by the students. They reacted positively to the teen-aged host and hostess but said that some of the games on the game segments were probably more suited to younger children. The segment which received the highest overall rating was the skit segments, mainly because of the humour and the quick action.

The analysis of the recall of segment answers indicated that most students caught the general drift of the action but missed important details.

IMPROVING EVALUATIONS: SOME GUIDELINES

A variety of experimental designs and procedures are needed in order to conduct a formative evaluation of a second language or indeed of any television project.

One of the most important factors is the expert evaluation in which people who qualified for the test review the program independent of students to determine the answers to questions such as the suitability of the vocabulary, the accents, the correspondence with the curriculum, the plot level, etc.

Another procedure involves the use of a few students to determine empirically the vocabulary level of the program in comparison to the level of the target audience, the attention of students in small groups and other such factors.

The field evaluation involves the actual in-class testing of the program, with teacher input and without, to determine the program's ability to achieve its intended objectives. This type of procecure was quite widely used in most evaluations.

Most evaluations which included field testing use experimental group, control group designs with or without some provision for randomization or statistical control of pre-treatment differences.

In evaluating second language materials, some or all of the factors mentioned below may have to be included. The factors are divided into three categories: factors which can be examined by viewing the program, factors which need viewers interacting with the program.

Factors which can be examined independently

1. Vocabulary level of the program

The vocabulary level of the program must be within the range of the students in order that they can benefit from the program. There are several ways of determining ideal vocabulary level. One way would be to compare the list of words used in the program with the list of words used in the instruction materials which are being used in the curriculum. Another means might be to pre-test the students on the vocabulary which is used in the program. The pre-test can take the form which is used in curriculum materials which usually consists of a series of pictures in which the student circles the correct alternative which is mentioned on an audio-tape. This type of test was used in both the "Vive La Compagnie" and the "Allons-Y" evaluations. A variation was used in the "Allons-Y" evaluation in which the student numbered one of twenty pictures with the number of the phrase or word. However, this test was a bit more difficult than the traditional test in which the student circled one of four alternatives.

2. The types of constructions used

The formats of constructions will effect comprehension of students. The formats of construction used in the program must match the formats which have been used by the students in their curriculum or which have been reinforced in the teaching. Once again, the types of constructions which the students understand can be tested or can be deduced from the program materials.

3. Accents

The accent of the actors must match those which the students are used to hearing. However, there is some discussion on this fact with one camp believing that the student must get used to hearing a variety of accents while the opposite view holds that the student must hear words with the same accent. This is proven to be an especially vexing problem in Canada where the French spoken in the various regions differs in terms of accent and also differs from French spoken in France. Similarly, the types of inflections placed on words must be similar to those heard by the students. One of the ways this task can be done is to ask expert personnel to judge the accents and inflections or to pre-test on a small number of students.

4. Image presented by the actors

In any materials which have as a goal the improvement of relationships between language groups, the image presented by the actors in the second language program must be very carefully monitored. Experts in the area must be asked to evaluate the image portrayed by the actors. If the actors or actresses exhibit a negative image to the students, this may well inhibit learning.

5. Contribution of production effects

The sets, and production techniques should be related to the objectives of the program. In addition, these factors should be similar to the kinds of television to which the students are accustomed. Makeshift sets are detected very quickly by students and cannot be used if the purpose is to present professional television. In addition, the sets should not in themselves distract from the objectives.

6. Correspondence of audio and visual message

The presentation of the new vocabulary words can and should be accompanied by their equivalent in a real referent. When the French or Spanish word for cup is used, the cup should be the most prominent object on the screen. Another method would be to use a visual cue, with perhaps one of the actors pointing to the cup, to indicate precisely which object is being referred to in the auditory message. In one of the programs, the weak correspondence between audio and visual stimuli meant that students did not clearly understand which objects an auditory message was referring to.

7. Speed of the actors' speech

The actors must speak at a pace which is correct for the students. If the pace is too quick the students will not understand the speech. However, if the speech is too slow, it seems unnatural and will distract the students in other ways.

8. Level of content

If a story is being used, the level of the plot must match the interest level of the students. If, as in "Allons-Y", a segment approach is used, the segments themselves must also match the interest of the students.

9. Complexity of plot

The plot must be complex enough to hold the students' interest and yet simple enough so that the students will understand it. Teachers and other professionals can judge the complexity of plot relative to a group, however, a better means might be to use small groups of students and pilot studies to determine the suitability of the plot to the intended audience.

10. Analysis of the program into information units

Using a procedure outlined by Friedlander (1974) the program should be broken into information units so that these units can be tested in comprehension tests or vocabulary tests later on. Analysis of information units would probably take place by teachers familiar with the program or by researchers who are conducting the evaluation.

Factors which must be examined in context

1. Correspondence with the curriculum

The television program which is used should correspond in some way to the curriculum which is being used in the schools. If the vocabulary constructions or structure of the program is contrary to that of the instructional programming used in the schools, a difficulty will emerge. Once again, this evaluation could be conducted by expert researchers or by teachers who are familiar with both the program and the instructional materials being used in the school. A subsidiary point is that the program should help fulfill some of the curriculum objectives in the curriculum. It is important that the objectives of the program be somewhat the same as the objectives of the overall curriculum.

2. How the program can be used

In many cases, the schedule of the junior high and high schools would make it impossible for the program to be used on an off-the-air basis unless cable scheduling can be arranged. The accessibility of the program to the target audience, then, must be clearly identified. In some of the Nova Scotia evaluations, it was discovered that because of a chain of circumstances, many of the program materials could simply not be used by a large part of the target audience due to the inaccessibility of equipment and the problem of rotating schedules.

3. Preparatory and follow-up work

Through interaction with teachers the ideal means of preparatory and follow-up work can and should be specified. This preparatory and follow-up work should help to achieve the program's objectives and the objectives of the overall curriculum. One of the only ways that this effect can be tested is to contrast the use of teacher input with a comparison group who do not have teacher input such as was done in the "Allons-Y" research.

4. Teacher knowledge about using the program

Through direct contact with the teachers, the evaluation should determine whether teachers are able to use the programs or whether concrete guidelines should be provided. By having the teacher view a number of programs and asking how the programs would be used, the evaluation team should be able to determine whether concrete guidelines are needed to use the program.

5. The preparation of students

The evaluation team should try to determine whether concrete preparation in the viewing of educational television materials is necessary for students. Certainly, they need no education to watch television but they may need some help in gaining the most from an educational television learning experience.

Factors which need students viewing the program

1. Attention

Attention, interest or appeal of the program, is one of the first variables to be tested when students are involved. Attention can be measured by various means ranging from a simple count of eyes on screen, to program analyzer techniques, conjugate reinforcement (Lindsley, O. R. 1962, Ogston 1975), eye movement studies (O'Bryan, 1975), distractor analysis (Palmer, 1974), or similar techniques. Lewis (1973) reviewed various means of measuring attention to audio-visual materials.

The purpose of measuring the attention is to ensure that each segment of the program is interesting enough to be watched by the students. If the student is not watching the program, all future measurement will be in jeopardy.

Attention can be tested either in the group situation or in a viewing situation in which one student or a small group watching the program is monitored by one of the techniques mentioned above to increase the vocabulary level of students, help them hear the spoken language and help them improve their ability to construct sentence in the target language. The purpose of the on-site evaluation is to determine whether any vocabulary or correct construction has been acquired through interaction with the program.

3. Comprehension of the Program

Often in the target language, although the vocabulary is at a level which can be understood, the complexity of the plot or other factors make it hard for students to understand the exact course of action. As a result, the comprehension of the program must be measured. There are various procedures to assess comprehension, the simplest one being an interview with the student to determine what he or she remembered from the program. Other means which could be used include a question technique with multiple alternatives, or a recall test in which students are simply asked to write down what happened.

4. Attitude to program segments

The enjoyment of the various segments or total program is also one of the most important areas to measure. In the case of second language instruction although the program may fail to teach any vocabulary or construction, or even in fact may fail to be completely understood by students, they may enjoy the interaction with the program. The attitude to program segments also would measure attitude to production variables, level of plot, plot complexity, the actors and any other relevant information.

5. Attitude to production

The students who view our second language programs will have viewed a great deal of television by the time they interact with one of the programs that we produce. As a result, it is important to determine whether they consider the production level to be similar to that of the commercial programs that they watch. If the production level is seen to be inferior to the programs they watch, this will likely interfere with their learning or enjoyment of the program. The attitude to production can be obtained using open ended statements or using Likert type scales.

6. Post-program monitoring

Many of the effects of the second language program will not be evident immediately after the program is viewed. However, they may well be exhibited later on after the students have had time to digest the program and fit it in to their regular work. The purpose of post-program activity monitoring is to determine the ability of the program to produce changes in the behavior of the students. The usage of vocabulary constructions learned in the program, the repetition of the types of activities produced in the program would seem to indicate whether the program would be successful in promoting long-term change. Post-program monitoring activity would have to take place through observational procedures or other similar devices.

7. Attitude to French course

One would expect that if one group of students is using a French television series while the other group is using standard curricular materials there might be a difference in attitudes. Gardner and Smythe (1975) have suggested an attitude to French course questionnaire which could probably be used to measure attitudes to the French course.

Conclusions

The formative evaluation of bilingual television programs requires a multi-faceted research approach so that the various aspects of the programs can be examined. No one approach, such as field testing, can provide guidance to producers of television programs. A variety of methods provide answers to the many questions. Many of the techniques mentioned above have been tried and tested, both the two evaluations described above and in other formative evaluations of television programs. However, some of the other procedures

have emanated from the research but have not been fully tested. There is much work to be done in this field by researchers in the field of educational technology. The best research designs and measurement techniques must be marshalled to advance the learning in this area.

REFERENCES

Fleming, W. G. Evaluation of four educational television program series produced by the Ontario Educational Communications Authority. Research Paper No. 26. Toronto: O.E.C.A. 1972.

Friedlander, D. Z. Maximizing comprehension of informational films and videotape: formative evaluation during production. Mimeo. New England Instructional Television Research Center, University of Hartford, Hartford, Connecticut. 1974.

Gardner, R. C., Ginsberg, R. E. and Smythe, P. C. Attitudes and motivation in second language learning: a social psychological perspective. The Canadian Modern Language Review. 1976 32 243ff.

Gardner, R. C. and Smythe, P. C. Second language acquisition: a social psychological approach. Research Bulletin 232, Department of Psychology, University of Western Ontario, London, Ontario. 1975.

Johnston-Doyle, K. and Lewis, R.F. A content analysis of student comments on "Le masque de Marie Louise": A program in the "Allons-Y" series. Halifax, NS: Atlantic Institute of Education R. & D. Bulletin No. 1, 1978.

Laosa, Luis, M. Formative research and evaluation in the application of television media to bilingual multicultural education. California Journal of Educational Research, 1975, 25 (5) 302-309.

Lewis, Richard F. An evaluation of "Allons-Y" Halifax, NS: Atlantic Institute of Education. Technical report, 1979.

Lewis, Richard F. The effect of stimulus color and degree of realism on attention. Syracuse University, Syracuse, New York. Unpublished doctoral dissertation, 1973.

Lewis, R. F. and Fisher, G. An evaluation of the junior high school series "Vive La Compagnie" Halifax, NS: Atlantic Institute of Education. Technical report, 1977.

Lindsley, O. R. A behavioral measure of television viewing. Journal of Advertising Research, 1962, 2 2-12.

O'Bryan, K. G. Television and eye movement research. Mimeo. Toronto: Ontario Educational Communications Authority, 1975.

Ogston, D. G. An instructional television behavioral pretest measure and its relationship to learning. Journal of Educational Technology Systems, 1975 4 (2), 125-238.

Palmer, E. L. Formative research in the production of television for children. Chapter 12 of the 73rd. Yearbook of the National Society for the Study of Education. Chicago, Illinois: National Society for the Study of Education, 1974.

Williams, F., Van Wart, G. and Stanford, M. National evaluation of a Spanish/English educational television series. Austin, Texas: Center for Communication Research, University of Texas, 1973.